Explanatory and Response Variables

Most statistical studies examine data on more than one variable. In many of these settings, the two variables play different roles.

Response Variable, Explanatory Variable

A **response variable** measures an outcome of a study.

An **explanatory variable** may help explain or influence changes in a response variable.

Note: In many studies, the goal is to show that changes in one or more explanatory variables actually *cause* changes in a response variable. However, other explanatory-response relationships don't involve direct causation.

Scatterplot

The most useful graph for displaying the relationship between two quantitative variables is a **scatterplot.**

A **scatterplot** shows the relationship between two quantitative variables measured on the same individuals.
The values of one variable appear on the horizontal axis, and the values of the other variable appear on the vertical axis.
Each individual in the data appears as a point on the graph.

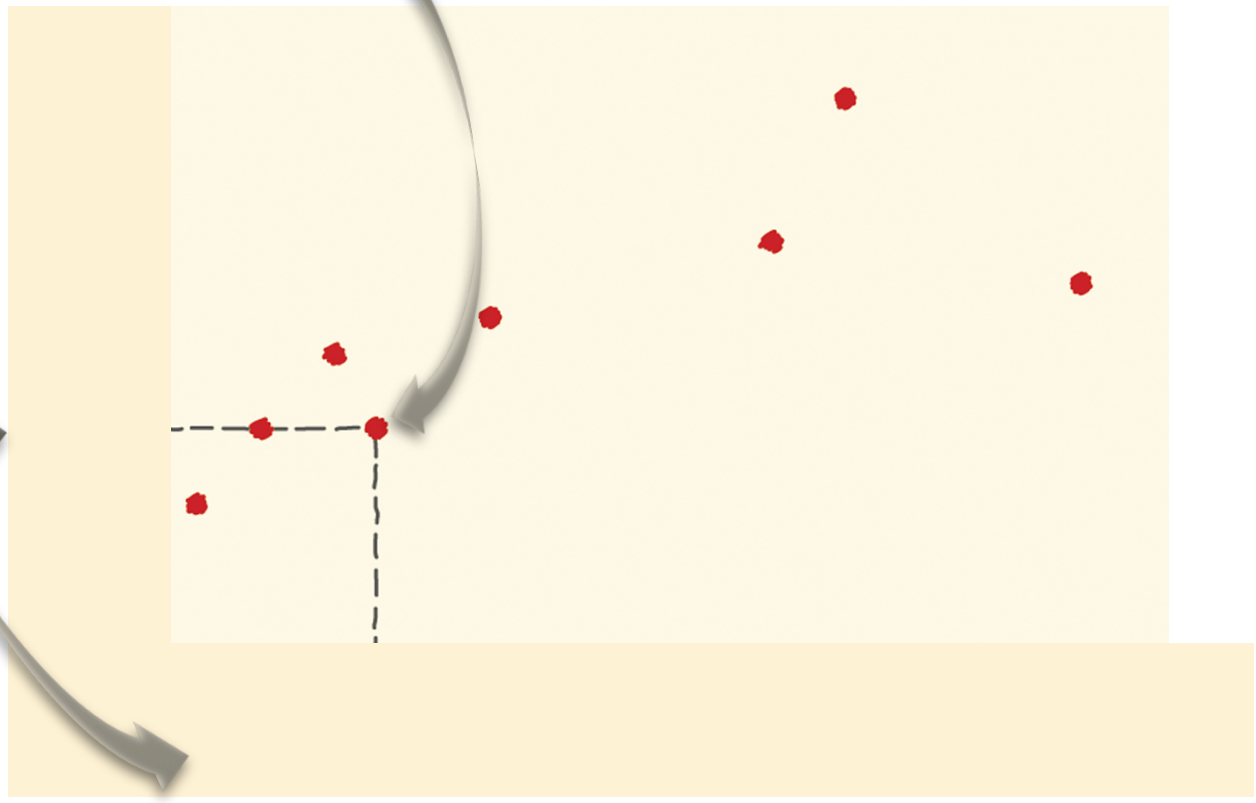
How to Make a Scatterplot

1. Decide which variable should go on each axis. If a distinction exists, plot the explanatory variable on the x-axis and the response variable on the y-axis.
2. Label and scale your axes.
3. Plot individual data values.

Scatterplot

Example: Make a scatterplot of the relationship between body weight and pack weight for a group of hikers.

Body weight (lb)	120	187	109	103	131	165	158	116
Backpack weight (lb)	26	30	26	24	29	35	31	28



Interpreting Scatterplots

To interpret a scatterplot, follow the basic strategy of data analysis from Chapters 1 and 2. Look for patterns and important departures from those patterns.

How to Examine a Scatterplot

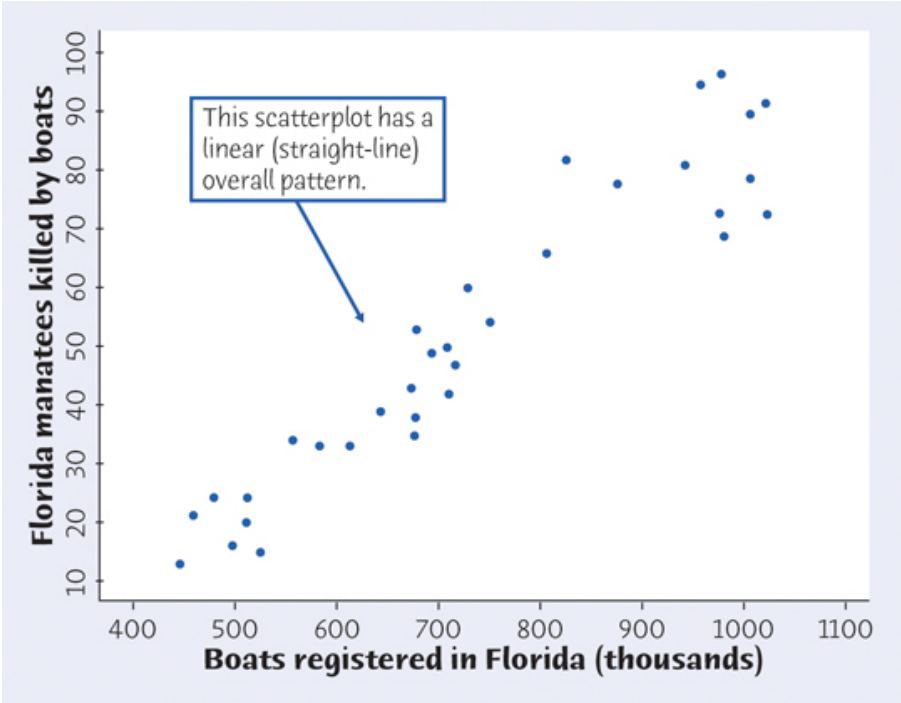
As in any graph of data, look for the *overall pattern* and for striking *departures* from that pattern.

- You can describe the overall pattern of a scatterplot by the **direction, form,** and **strength** of the relationship.
- An important kind of departure is an **outlier,** an individual value that falls outside the overall pattern of the relationship.

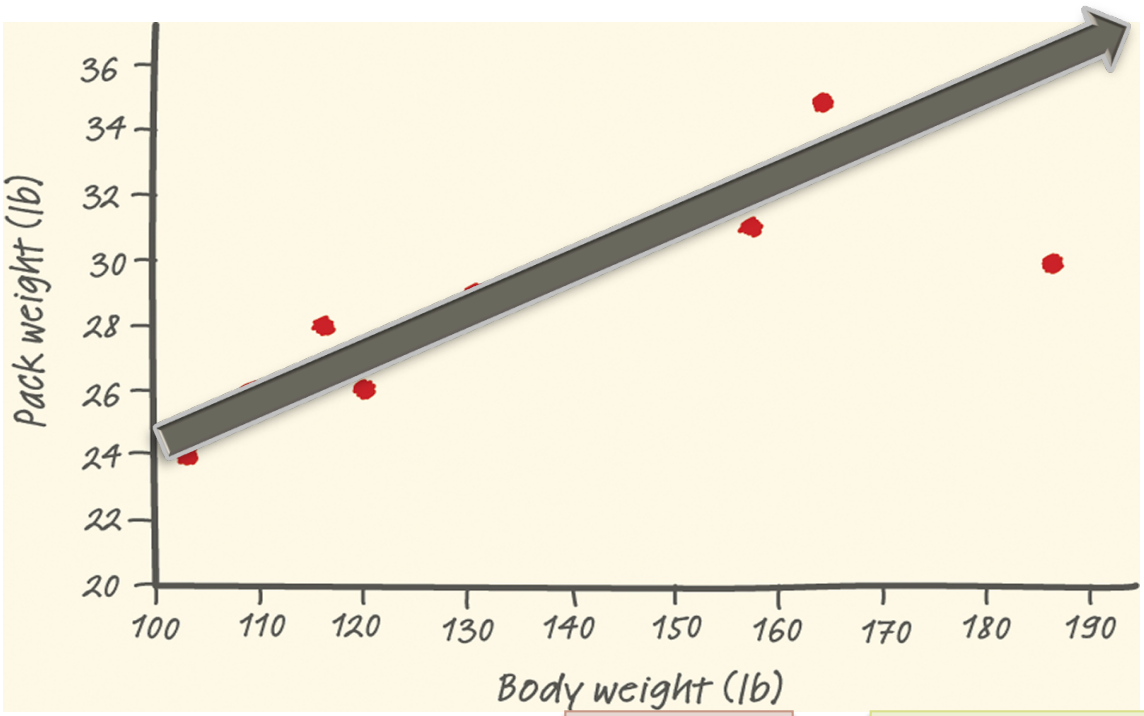
Interpreting Scatterplots

Two variables have a **positive association** when above-average values of one tend to accompany above-average values of the other, and when below-average values also tend to occur together.

Two variables have a **negative association** when above-average values of one tend to accompany below-average values of the other.



Interpreting Scatterplots



Outlier

There is one possible outlier: the hiker with the body weight of 187 pounds seems to be carrying relatively less weight than are the other group members.

Strength **Direction** **Form**

- ✓ There is a moderately strong, positive, linear relationship between body weight and pack weight.
- ✓ It appears that lighter hikers are carrying lighter backpacks.

Measuring Linear Association

- A scatterplot displays the strength, direction, and form of the relationship between two quantitative variables.

The **correlation *r*** measures the strength of the linear relationship between two quantitative variables.

$$r = \frac{1}{n-1}\sum\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

- *r* is always a number between -1 and 1.
- $r > 0$ indicates a positive association.
- $r < 0$ indicates a negative association.
- Values of *r* near 0 indicate a very weak linear relationship.
- The strength of the linear relationship increases as *r* moves away from 0 toward -1 or 1.
- The extreme values $r = -1$ and $r = 1$ occur only in the case of a perfect linear relationship.

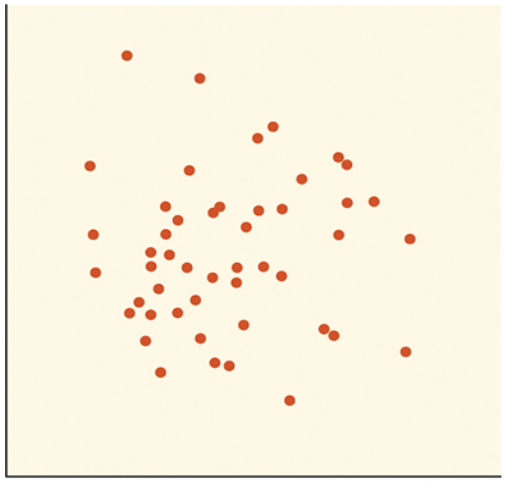
Facts About Correlation

1. Correlation makes no distinction between explanatory and response variables.
2. *r* has no units and does not change when we change the units of measurement of *x*, *y*, or both.
3. Positive *r* indicates positive association between the variables, and negative *r* indicates negative association.
4. The correlation *r* is always a number between -1 and 1.

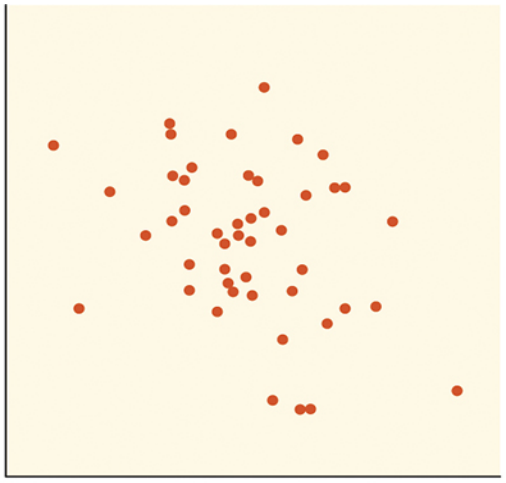
Cautions:

- Correlation requires that both variables be quantitative.
- Correlation does not describe curved relationships between variables, no matter how strong the relationship is.
- Correlation is not resistant. *r* is strongly affected by a few outlying observations.
- Correlation is not a complete summary of two-variable data.

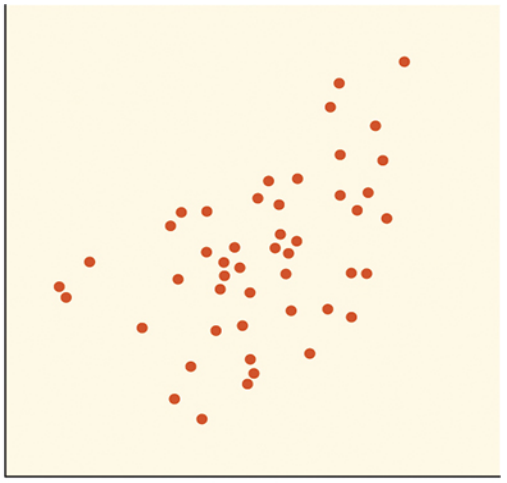
Correlation



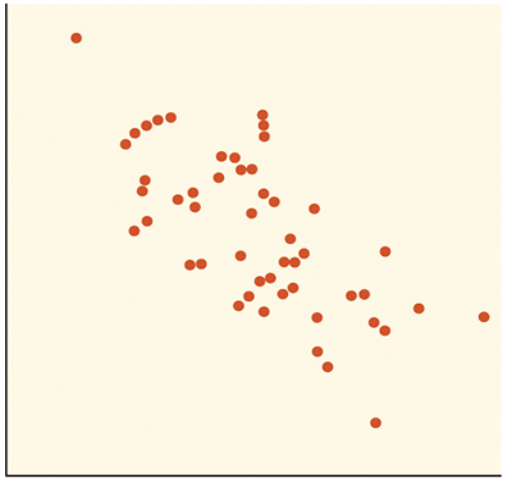
Correlation $r = 0$



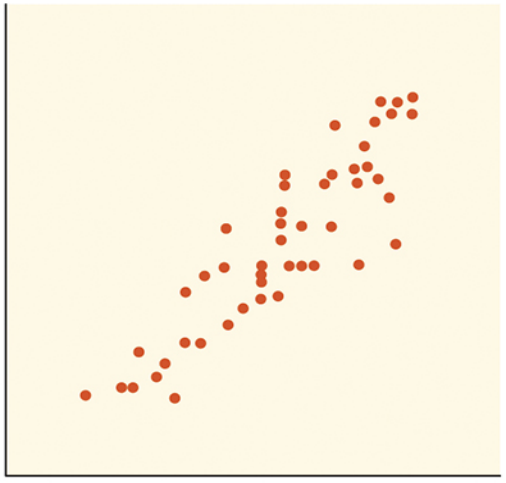
Correlation $r = -0.3$



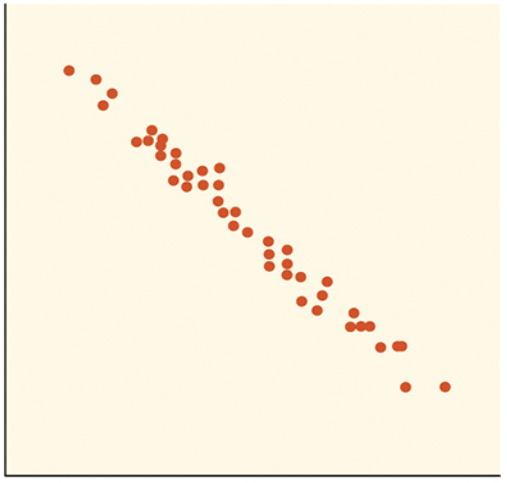
Correlation $r = 0.5$



Correlation $r = -0.7$



Correlation $r = 0.9$

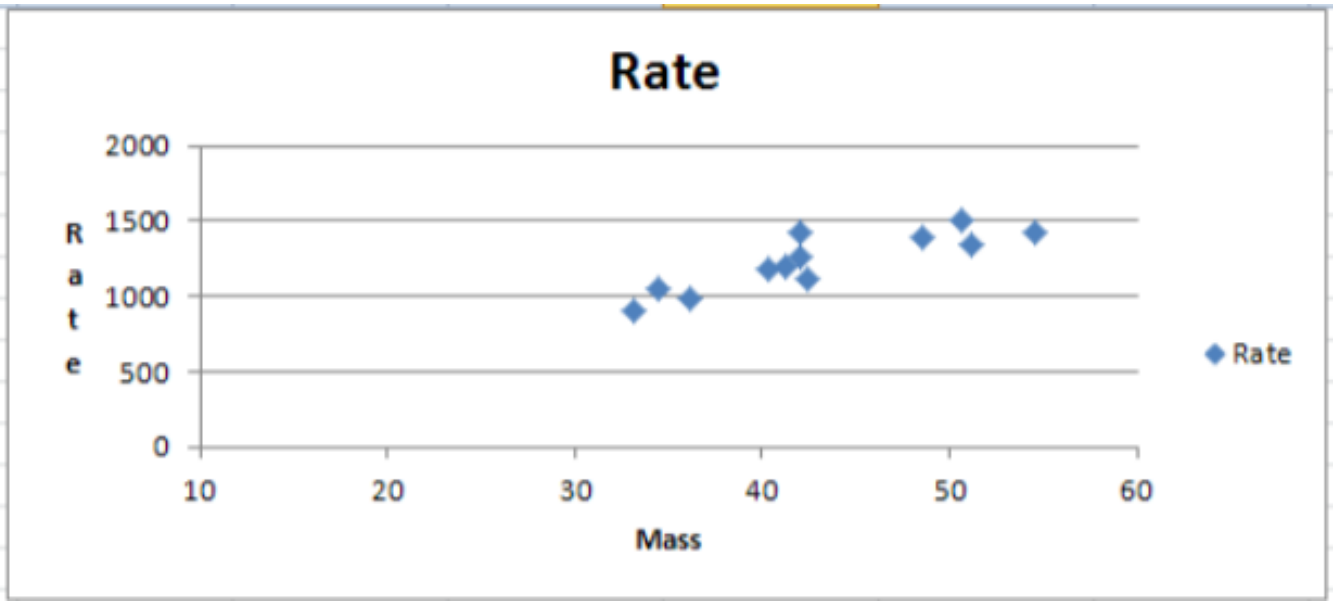


Correlation $r = -0.99$

Ex. 4.5 Do heavier people burn more energy?

We have data on the lean body mass and resting metabolic rate for 12 women who are subjects in a study of dieting. Lean body mass, given in kilograms, is a person's weight leaving out all fat. Metabolic rate is measured in calories burned per 24 hours.

Mass	Rate
36.1	995
54.6	1425
48.5	1396
42	1418
50.6	1502
42	1256
40.3	1189
33.1	913
42.4	1124
34.5	1052
51.1	1347
41.2	1204



The scatterplot shows a positive direction, linear form, and moderately strong association.